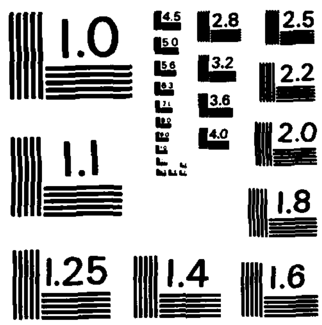
AD-A147 614 EMPIRICAL BAYES RULES FOR SELECTING GOOD BINOMIAL POPULATIONS(U) PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS S S GUPTA ET AL. OCT 84 TR-84-37 1/1

UNCLASSIFIED N00014-84-C-0167 F/G 12/1 NL

END


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

Good Binomial Populations

by

Shanti S. Gupta and Ta Chen Liang
Purdue University

Technical Report #84-37

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

Empirical Bayes Rules for Selecting
Good Binomial Populations

by

Shanti S. Gupta and Ta Chen Liang
Purdue University

Technical Report #84-37

Purdue University
Department of Statistics

October 1984

*This research was supported by the Office of Naval Research Contract N00014-84-0167 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Empirical Bayes Rules for Selecting Good Binomial Populations

1. Introduction

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem in the sequence as Bayes decision problems with respect to an unknown prior distribution on the parameter space and then use the accumulated observations to improve the decision rule at each stage. This approach is due to Robbins (1955, 1964). Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the nth decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was known and the Bayes rule with respect to this prior distribution was used.

Empirical Bayes rules have been derived for multiple decision problems by Deely (1965). He considered selecting a subset containing the best population. Van Ryzin (1970), Huang (1975), Van Ryzin and Susarla (1977) and Singh (1977) also studied some multiple decision problems by using empirical Bayes approach. Recently, Gupta and Hsiao (1983) studied some empirical Bayes rules for selecting good populations with respect to a standard or a control. In their paper, the underlying population π_i is uniformly distributed with parameter θ_i, $i = 0,1,\ldots,k$, and π_0 is a control population. π_i is said to be good if $\theta_i \geq \theta_0$ and to be bad if $\theta_i < \theta_0$. Let $a \subset \{1,\ldots,k\}$ be an action. When action a is taken, it means that π_i is selected as good if $i \in a$, and excluded as bad if $i \notin a$. With the loss

function $L(\underset{\sim}{\theta},a) = \sum_{j\in a} (\theta_0-\theta_j)I_{(0,\theta_0)}(\theta_j) + \sum_{j\notin a}(\theta_j-\theta_0)\ I_{(\theta_0,\infty)}(\theta_j)$, where $\underset{\sim}{\theta} = (\theta_0,\theta_1,\ldots,\theta_k)$, they proposed some empirical Bayes rules for the problem of selecting good populations with respect to a standard or a control.

For a similar problem, if the underlying populations have binomial distribution, then, in general, it is impossible to find a sequence of empirical Bayes rules which is asymptotically optimal in the sense mentioned above (see Robbins (1964), Samuel (1963) and Singh (1977)). In this paper, we are concerned with this problem. Two cases have been studied: one is that the prior distribution is completely unknown and the other is that the prior distribution is symmetrical about $p = \frac{1}{2}$, but its form is still unknown. In each case, empirical Bayes rules are derived and the rate of convergence of corresponding empirical Bayes rules is also studied. In each case, it is found that the order of the rate of convergence is $O(\exp(-c_i n))$ for some $c_i > 0$, $i = 1,2$. For the case when the prior distribution is symmetrical about $p = \frac{1}{2}$, in order to improve the performance of the sequence of empirical Bayes rules, two smoothing methods are studied. Some Monte Carlo studies have also been carried out. The results indicate that the smoothed competitors actually perform better than the original empirical Bayes rules.

2. Formulation of the Empirical Bayes Approach

Let $\pi_0, \pi_1,\ldots,\pi_k$ denote $k + 1$ populations and let X_i be a random observation from π_i. Assume that $X_i \sim B(N_i,p_i)$, where $p_i \in (0,1)$ and N_i is fixed and known. Let π_0 be the control population. For each $i = 1,\ldots,k$, population π_i is said to be good if $p_i \geq p_0$ and to be bad if $p_i < p_0$, where the control parameter p_0 is either known or unknown. Our goal is to derive some empirical Bayes rules to select all the good populations and exclude all the bad populations.

When the control parameter p_0 is known, the empirical Bayes framework can be formulated as follows:

(1) Let $\Omega = \{\underset{\sim}{p} | \underset{\sim}{p} = (p_1, \ldots, p_k),\ p_i \in (0,1)$ for $i = 1,2,\ldots,k\}$. For each $\underset{\sim}{p} \in \Omega$, define $A(\underset{\sim}{p}) = \{i | p_i \geq p_0\}$, $B(\underset{\sim}{p}) = \{i | p_i < p_0\}$. That is, $A(\underset{\sim}{p})(B(\underset{\sim}{p}))$ is the set of indices of good (bad) populations.

(2) Let $A = \{a | a \subset \{1,2,\ldots,k\}\}$ be the action space. When action a is taken, it means that population π_i is selected as a good population if $i \in a$, and excluded as a bad population if $i \notin a$.

(3) The loss function $L(\underset{\sim}{p}, a)$ is defined as follows:

$$L(\underset{\sim}{p}, a) = \sum_{i \in A(\underset{\sim}{p}) - a} (p_i - p_0) + \sum_{i \in a - A(\underset{\sim}{p})} (p_0 - p_i) \tag{2.1}$$

where the first summation is the loss due to not selecting some good populations and the second summation is the loss due to selecting some bad populations.

(4) Let $dG(\underset{\sim}{p}) = \prod_{i=1}^{k} dG_i(p_i)$ be the prior distribution over the parameter space Ω, where $G_i(\cdot)$ are unknown for all $i = 1,2,\ldots,k$.

(5) For each i, let (X_{ij}, P_{ij}), $i = 1,2,\ldots$, be pairs of random variables associated with population π_i, where X_{ij} is observable but P_{ij} is not observable. P_{ij} has distribution G_i. Conditional on $P_{ij} = p_{ij}$, X_{ij} is binomially distributed with parameters N_i and p_{ij}. Some additional observations $\underset{\sim}{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijn_i})$ are also available. Conditional on $P_{ij} = p_{ij}$, X_{ij} and Y_{ijm}, $m = 1,\ldots,n_i$, are i.i.d. The jth stage observations are denoted by $\underset{\sim}{Z}_j$. That is, $\underset{\sim}{Z}_j = ((X_{1j}, \underset{\sim}{Y}_{1j}), \ldots, (X_{kj}, \underset{\sim}{Y}_{kj}))$.

(6) Let $\underset{\sim}{X} = (X_1,\ldots,X_k)$ be the present observation. Conditional on $\underset{\sim}{p} = (p_1,\ldots,p_k)$, $\underset{\sim}{X}$ has probability function

$$f(\underset{\sim}{x}|\underset{\sim}{p}) = \prod_{i=1}^{k} f_i(x_i|p_i) = \prod_{i=1}^{k} \binom{N_i}{x_i} p_i^{x_i} (1-p_i)^{N_i-x_i}.$$

Finally, since we are interested in Bayes rule, we can restrict our attention to the nonrandomized rules.

(7) Let $D = \{d|d : \chi \to A$, being measurable$\}$ be the set of nonrandomized rules, where $\chi = \prod_{i=1}^{k} \{0,1,\ldots,N_i\}$. For each $d \in D$, let $r(G,d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G,d)$ is the minimum Bayes risk.

When the control parameter p_0 is unknown, for the related framework, the indices in the associated notations should begin at 0 instead of at 1. In the sequel, (0) will be used to show this additional fact.

We now consider decision rules $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$ whose form depends on $\underset{\sim}{x}$ and $\underset{\sim}{Z}_j$, $j = 1,\ldots,n$. Let $r(G,d_n)$ be the Bayes risk associated with decision rule $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$. That is,

$$r(G,d_n) \equiv \sum_{\underset{\sim}{x} \in \chi} E \int_{\Omega} L(\underset{\sim}{p}, d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n))\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p})$$

where the expectation E is taken with respect to $(\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$. For simplicity, $d_n(\underset{\sim}{x}, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n)$ will be denoted by $d_n(\underset{\sim}{x})$.

Definition 2.1. A sequence of decision rules $\{d_n(\underset{\sim}{x})\}_{n=1}^{\infty}$ is said to be asymptotically optimal (a.o.) relative to the prior distribution G if $r(G,d_n) \to r(G)$ as $n \to \infty$.

For constructing a sequence of a.o. rules, we first need to find the minimum Bayes risk and the associated Bayes rule, say d_G. From (2.1), the Bayes risk associated with decision rule d is

$$r(G,d) = \sum_{\underset{\sim}{x}\in\mathcal{X}} \sum_{i\in d(\underset{\sim}{x})} \int_\Omega (p_0 - p_i)\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p})$$

$$+ \sum_{\underset{\sim}{x}\in\mathcal{X}} \sum_{i=1}^{k} \int_\Omega (p_i - p_0) I_{(p_0,1)}(p_i) f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}), \tag{2.2}$$

where $I_A(\cdot)$ is the indicator function of set A.

The second term in the right-hand side of (2.2) is a constant and does not affect the determination of the Bayes rule.

Let $\phi_{iG}(\underset{\sim}{x}) = \int_\Omega (p_0 - p_i) f(\underset{\sim}{x}|\underset{\sim}{p}) dG(\underset{\sim}{p})$. After integration, one obtains

$$\phi_{iG}(\underset{\sim}{x}) = \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1\\ j\neq i}}^{k} f_j(x_j)$$

where

$$\Delta_{iG}(\underset{\sim}{x}) = \begin{cases} p_0 f_i(x_i) - W_i(x_i) & \text{if } p_0 \text{ is known;} \\ W_0(x_0) f_i(x_i) - W_i(x_i) f_0(x_0) & \text{if } p_0 \text{ is unknown;} \end{cases} \tag{2.3}$$

$$f_i(x) = \int_0^1 f_i(x|p)\, dG_i(p) \quad \text{and}$$

$$W_i(x) = \int_0^1 p f_i(x|p) dG_i(p) = \int_0^1 \binom{N_i}{x} p^{x+1} (1-p)^{N_i - x} dG_i(p).$$

Since $f_i(x)$, the marginal probability function of X_i, is always positive for all $x = 0,1,\ldots,N_i$, $i = (0), 1,\ldots,k$, the Bayes rule d_G can be obtained as follows:

$$d_G(\underset{\sim}{x}) = \{i \mid \Delta_{iG}(\underset{\sim}{x}) \leq 0\}. \tag{2.4}$$

Now, for each $i = (0), 1,\ldots,k$, and for each $n = 1,2,\ldots$, let $W_{in}(x_i) \equiv W_{in}(x_i;\ (X_{i1},\underset{\sim}{Y}_{i1}),\ldots,(X_{in},\underset{\sim}{Y}_{in}))$ be an estimator of $W_i(x_i)$ and $f_{in}(x_i) \equiv f_{in}(x_i;\ (X_{i1},\underset{\sim}{Y}_{i1}),\ldots,(X_{in},\underset{\sim}{Y}_{in}))$ be an estimator of $f_i(x_i)$. Define

$$\Delta_{in}(\underset{\sim}{x}) = \begin{cases} W_{0n}(x_0)f_{in}(x) - W_{in}(x_i)f_{0n}(x_0) & \text{if } p_0 \text{ is unknown;} \\ \\ p_0 f_{in}(x_i) - W_{in}(x_i) & \text{if } p_0 \text{ is known;} \end{cases} \tag{2.5}$$

and

$$d_n(\underset{\sim}{x}) = \{i \mid \Delta_{in}(\underset{\sim}{x}) \leq 0\}. \tag{2.6}$$

If $W_{in}(x) \xrightarrow{p} W_i(x)$ and $f_{in}(x) \xrightarrow{p} f_i(x)$ for all $x = 0,1,\ldots,N_i$ where " $\xrightarrow{p}$ " means convergence in probability, then $\Delta_{in}(\underset{\sim}{x}) \xrightarrow{p} \Delta_{iG}(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in \chi$. Therefore, from a corollary of Robbins (1964), it follows that $r(G,d_n) \to r(G)$ as $n \to \infty$. So, the sequence of decision rules $\{d_n(\underset{\sim}{x})\}$ defined in (2.6) is asymptotically optimal for our selection problem. Hence, in the following, all we have to do is to find sequences of estimators, say $\{W_{in}(x)\}$ and $\{f_{in}(x)\}$, $i = (0),1,\ldots,k$, satisfying $W_{in}(x) \xrightarrow{p} W_i(x)$ and $f_{in}(x) \xrightarrow{p} f_i(x)$ for all $x = 0,1,\ldots,N_i$.

3. Case when the Prior Distribution is Completely Unknown

Robbins (1964) and Samuel (1963), respectively, pointed out that there was no way of approximating $W_i(x)$ just by using the observations $(X_{i1},\ldots,X_{in})$. In order to remedy this deficiency, we take, at each stage, some more observations $(Y_{ij1},\ldots,Y_{ijn_i})$ in our model where n_i can be any positive integer. For simplicity, let $n_i = 1$ for all $i = (0),1,\ldots,k$.

Estimation of $W_i(x)$ and $f_i(x)$

A usual estimator of $f_i(x)$ can be given as follows:

$$f_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_{ij}) \quad \text{for } x = 0,1,\ldots,N_i. \tag{3.1}$$

Then $f_{in}(x)$ is an unbiased estimator of $f_i(x)$, and by the strong law of large numbers, $f_{in}(x) \to f_i(x)$ with probability 1 for each $x = 0,1,\ldots,N_i$. Hence, $f_{in}(x) \xrightarrow{p} f_i(x)$ for all $x = 0,1,\ldots,N_i$.

For the estimation of $W_i(x)$, we consider the following. Define

$$V_{ij}(x) = Y_{ij}\, I_{\{x\}}(X_{ij}). \tag{3.2}$$

Under the condition (5) of Section 2, it is easy to see that $E[V_{ij}(x)] = N_iW_i(x)$. We then define

$$W_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} V_{ij}(x)/N_i. \tag{3.3}$$

Since $V_{ij}(x)$, $i = 1,2,\ldots$, are i.i.d. and bound, it is easy to show that $W_{in}(x) \to W_i(x)$ with probability one for all $x = 0,1,\ldots,N_i$. Now, let $\Delta_{in}(\underset{\sim}{x})$ and $d_n(\underset{\sim}{x})$ be defined in (2.5) and (2.6), respectively. From the discussion

of Section 2 and the construction of the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ through (2.5), (2.6), (3.1) and (3.3), we get the following result.

Theorem 3.1. For our decision problem, the sequence of decision rules $\{d_n\}_{n=1}^{\infty}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d_n\}$

Let $\{\delta_n\}_{n=1}^{\infty}$ be a sequence of empirical Bayes rules relative to the prior distribution G. Since the Bayes rule d_G achieves the minimum Bayes risk $r(G)$ relative to G, $r(G,\delta_n) - r(G) \geq 0$ for all $n = 1,2,\ldots$. Thus, the nonnegative difference $r(G,\delta_n) - r(G)$ is used as a measure of the optimality of the sequence of empirical Bayes rules $\{\delta_n\}_{n=1}^{\infty}$.

Definition 3.1. The sequence of empirical Bayes rules $\{\delta_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to G if $r(G,\delta_n) - r(G) \leq O(\alpha_n)$ as $n \to \infty$ where $\lim\limits_{n\to\infty} \alpha_n = 0$.

For each $i = 1,\ldots,k$, define $S_i = \{\underset{\sim}{x}\varepsilon\chi \mid \Delta_{iG}(\underset{\sim}{x}) < 0\}$, $T_i = \{\underset{\sim}{x}\varepsilon\chi \mid \Delta_{iG}(\underset{\sim}{x}) > 0\}$. Let $\varepsilon_1 = \min\limits_{\substack{\underset{\sim}{x}\varepsilon S_i \\ 1\leq i\leq k}} (-\Delta_{iG}(\underset{\sim}{x}))$, $\varepsilon_2 = \min\limits_{\substack{\underset{\sim}{x}\varepsilon T_i \\ 1\leq i\leq k}} (\Delta_{iG}(\underset{\sim}{x}))$ and $\varepsilon = \min(\varepsilon_1,\varepsilon_2)$. Since χ is a finite space, therefore $\varepsilon > 0$. Now,

$$
\begin{aligned}
0 &\leq r(G,d_n) - r(G) \\
&= \sum_{\underset{\sim}{x}\varepsilon\chi} \left\{ E\left[\sum_{i\varepsilon d_n(\underset{\sim}{x})} \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1\\ j\neq i}}^{k} f_j(x_j)\right] - \sum_{i\varepsilon d_G(\underset{\sim}{x})} \Delta_{iG}(\underset{\sim}{x}) \prod_{\substack{j=1\\ j\neq i}}^{k} f_j(x_j) \right\} \\
&= \sum_{i=1}^{k} \sum_{\underset{\sim}{x}\varepsilon S_i} (-1)\Delta_{iG}(\underset{\sim}{x}) P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \prod_{\substack{j=1\\ j\neq i}}^{k} f_j(x_j)
\end{aligned}
$$

$$+ \sum_{i=1}^{k} \sum_{\underset{\sim}{x} \varepsilon T_i} \Delta_{iG}(\underset{\sim}{x}) P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\} \prod_{\substack{j=1 \\ j \neq i}}^{k} f_j(x_j)$$

$$\leq \sum_{i=1}^{k} \left\{ \sum_{\underset{\sim}{x} \varepsilon S_i} P\{\Delta_{in}(\underset{\sim}{x}) > 0\} + \sum_{\underset{\sim}{x} \varepsilon T_i} P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\} \right\} \tag{3.4}$$

where the last inequality is due to the fact that $0 < f_j(x_j) \leq 1$ and $|\Delta_{iG}(\underset{\sim}{x})| \leq 1$. From (3.4), it suffices to consider the behavior of $P\{\Delta_{in}(\underset{\sim}{x}) > 0\}$ when $\underset{\sim}{x} \varepsilon S_i$ and that of $P\{\Delta_{in}(\underset{\sim}{x}) \leq 0\}$ when $\underset{\sim}{x} \varepsilon T_i$ as $n \to \infty$ for each $i = 1,2,\ldots,k$.

For each $\underset{\sim}{x} \varepsilon S_i$,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} = P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > -\Delta_{iG}(\underset{\sim}{x})\}$$

$$\leq P\{\Delta_{in}(\underset{\sim}{x}) - \Delta_{iG}(\underset{\sim}{x}) > \varepsilon\}$$

When p_0 is known,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\} \leq P\{p_0 f_{in}(x_i) - p_0 f_i(x_i) - W_{in}(x_i) + W_i(x_i) > \varepsilon\}$$

$$\leq P\{f_{in}(x_i) - f_i(x_i) > \tfrac{\varepsilon}{2}\} + P\{W_{in}(x_i) - W_i(x_i) < -\tfrac{\varepsilon}{2}\} \tag{3.5}$$

since $p_0 \varepsilon (0,1)$.

When p_0 is known,

$$P\{\Delta_{in}(\underset{\sim}{x}) > 0\}$$

$$\leq P\{W_{0n}(x_0)f_{in}(x_i) - W_{in}(x_i)f_{0n}(x_0) - W_0(x_0)f_i(x_i) + W_i(x_i)f_0(x_0) > \varepsilon\}$$

$$\leq P\{f_{in}(x_i)[W_{0n}(x_0) - W_0(x_0)] > \tfrac{\varepsilon}{4}\} + P\{W_0(x_0)[f_{in}(x_i) - f_i(x_i)] > \tfrac{\varepsilon}{4}\}$$

$$+ P\{f_{0n}(x_0)[W_{in}(x_i) - W_i(x_i)] < -\tfrac{\varepsilon}{4}\} + P\{W_i(x_i)[f_{0n}(x_0) - f_0(x_0)] < -\tfrac{\varepsilon}{4}\}$$

$$\leq P\{W_{0n}(x_0) - W_0(x_0) > \tfrac{\varepsilon}{4}\} + P\{f_{in}(x_i) - f_i(x_i) > \tfrac{\varepsilon}{4}\}$$

$$+ P\{W_{in}(x_i) - W_i(x_i) < -\tfrac{\varepsilon}{4}\} + P\{f_{0n}(x_0) - f_0(x_0) < -\tfrac{\varepsilon}{4}\}. \tag{3.6}$$

In (3.6), the last inequality is due to the fact that $0 \leq W_i(x_i), f_i(x_i) \leq 1$ and that $0 \leq W_{in}(x_i), f_{in}(x_i) \leq 1$ where the latter can be easily checked from (3.1), (3.2) and (3.3).

(3.5) and (3.6) show that it suffices to consider the behavior of $P\{|f_{in}(x_i) - f_i(x_i)| > \delta\}$ and $P\{|W_{in}(x_i) - W_i(x_i)| > \delta\}$ for some $\delta > 0$.

From (3.2) and (3.3), $W_{in}(x)-W_i(x) = \sum_{j=1}^{n} A_{ij}(x)/n$ where $A_{ij}(x) = Y_{ij}I_{\{x\}}(X_{ij})/N_i-W_i(x)$. It is easy to see that $A_{ij}(x)$, $j = 1,\ldots,n$, are i.i.d. with mean 0 and finite variance, say $\beta_i(x)$, since $|A_{ij}(x)| \leq 1$. Therefore, for $m \geq 2$,

$$E[A_{ij}^m(x)] \leq E[|A_{ij}(x)|^m] \leq E[|A_{ij}(x)|^2] = \beta_i(x) \leq \tfrac{1}{2}\beta_i(x)m!.$$

Let $B_n(x) = n\beta_i(x)$. Thus, by Bernstein's inequality (see Ibragimov and Linnik (1971), page 169), for any $\delta > 0$,

$$\begin{aligned} &P\{|W_{in}(x) - W_i(x)| > \delta\} \\ &= P\{|\sum_{j=1}^{n} A_{ij}(x)| > n^{\frac{1}{2}}\delta\beta_i^{-\frac{1}{2}}(x)B_n^{\frac{1}{2}}(x)\} \qquad (3.7) \\ &\leq P\{|\sum_{j=1}^{n} A_{ij}(x)| > 2B_n^{\frac{1}{2}}(x)\min(\tfrac{1}{2}n^{\frac{1}{2}}\delta\beta_i^{-\frac{1}{2}}(x),\ \tfrac{1}{2}n^{\frac{1}{2}}\beta_i^{\frac{1}{2}}(x))\} \\ &\leq 2\exp\{-\frac{n}{4}\min(\delta^2\beta_i^{-1}(x),\ \beta_i(x))\}. \end{aligned}$$

Similarly, from (3.1), $f_{in}(x) - f_i(x) = \sum_{j=1}^{n} C_{ij}(x)/n$ where $C_{ij}(x) = I_{\{x\}}(X_{ij}) - f_i(x)$. Also, $C_{ij}(x)$, $j = 1,\ldots,n$, are i.i.d. with mean 0 and $|C_{ij}(x)| \leq 1$ and hence with finite variance, say $\alpha_i(x)$. Applying Bernstein's inequality again, we obtain

$$P\{|f_{in}(x) - f_i(x)| > \delta\} \leq 2\exp\{-\frac{n}{4}\min(\delta^2\alpha_i^{-1}(x),\ \alpha_i(x))\}. \qquad (3.8)$$

Thus, we take $\delta = \frac{\varepsilon}{4}$ if p_0 is unknown or take $\delta = \frac{\varepsilon}{2}$ if p_0 is known. Then, from (3.5) $\sim$ (3.8), for each $\underset{\sim}{x} \in S_i$,

$$\begin{aligned} P\{\Delta_{in}(\underset{\sim}{x}) > 0\} &\leq O(\exp\{-\frac{n}{4}\min(\delta^2\alpha_i^{-1}(x_i),\ \alpha_i(x_i))\}) \\ &+ O(\exp\{-\frac{n}{4}\min(\delta^2\beta_i^{-1}(x_i),\ \beta_i(x_i))\}). \end{aligned} \qquad (3.9)$$

Following an argument analogous to the above, we also get the conclusion given below:

For each $\underset{\sim}{x} \epsilon T_i$, $i = 1,\ldots,k$,

$$P\{\Delta_{in}(\underset{\sim}{x}) \le 0\} \le O(\exp\{-\frac{n}{4}\min(\delta^2\alpha_i^{-1}(x_i),\ \alpha_i(x_i))\})$$
$$+\ O(\exp\{-\frac{n}{4}\min(\delta^2\beta_i^{-1}(x_i),\ \beta_i(x_i))\}). \qquad (3.10)$$

Now, let $c_1 = \frac{1}{4}\min(b_1,b_2)$ where $b_1 = \min\limits_{m\le i\le k}\left[\min\limits_{0\le x\le N_i}(\delta^2\alpha_i^{-1}(x),\ \alpha_i(x))\right]$,

$b_2 = \min\limits_{m\le i\le k}\left[\min\limits_{0\le x\le N_i}(\delta^2\beta_i^{-1}(x),\ \beta_i(x))\right]$, here $m = 1$ if p_0 is known and

$m = 0$ if p_0 is unknown. It is clear that $c_1 > 0$ since $\beta_i(x) > 0$, $\alpha_i(x) > 0$ and χ is finite. Thus, we have the following theorem:

<u>Theorem 3.2</u>. Let $\{d_n\}_{n=1}^{\infty}$ be the sequence of asymptotically optimal rules described in Theorem 3.1. Then, $r(G,d_n)-r(G) \le O(\exp\{-c_1 n\})$ for some $c_1 > 0$.

4. <u>Case when $G_i(\cdot)$ are Symmetrical about $p = 1/2$</u>

In this section, we suppose that there is sufficient information to tell us that $G_i(\cdot)$ are symmetrical about $p = 1/2$ for all $i = (0), 1,\ldots,k$. Further, we also assume that N_i are even integers for all $i = (0), 1,\ldots,k$.

<u>Estimation of $W_i(x)$ and $f_i(x)$</u>

Under the above assumptions, $f_i(x) = f_i(N_i - x)$ for all $x = 0,1,\ldots,N_i$.

Therefore, it is reasonable to use

$$f_{in}^{1}(x) \equiv f_{in}^{1}(N_i-x) = \begin{cases} \dfrac{1}{2n} \displaystyle\sum_{j=1}^{n} I_{\{x,N_i-x\}}(X_{ij}) & \text{for } x \neq \dfrac{N_i}{2}, \\ \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} I_{\{x\}}(X_{ij}) & \text{for } x = \dfrac{N_i}{2} \end{cases} \tag{4.1}$$

to estimate $f_i(x)$.

For $W_i(x)$, $x = 0,1,\ldots,N_i$ we will construct a sequence of consistent estimators $\{W_{in}^{1}(x)\}$, in terms of $f_{in}^{1}(y)$, $y \in \{0,1,\ldots,N_i\}$, by using the observations $(X_{ij}, j = 1,\ldots,n)$ only. The following lemma is very helpful for the above purpose.

Lemma 4.1. Suppose that the prior distribution $G_i(\cdot)$ is symmetric about $p = 1/2$. Then

(a) $W_i(x) = \frac{x+1}{N_i-x} W_i(N_i-x-1)$ for each $x = 0,1,\ldots,N_i - 1$.

(b) $W_i(x) + W_i(N_i-x) = f_i(x) = f_i(N_i-x)$ for each $x = 0,1,\ldots,N_i$.

(c) Furthermore, if N_i is an even integer, then, $W_i\left(\frac{N_i}{2}\right) = \frac{1}{2} f_i\left(\frac{N_i}{2}\right)$.

Proof: Direct computation.

Theorem 4.1. Suppose that $G_i(\cdot)$ is symmetric about $p = 1/2$ and N_i is an even integer. Then, for each $x = 0,1,\ldots,N_i$, $W_i(x)$ can be represented as a linear function of $f_i(y)$, $y = 0,1,\ldots,N_i$.

Proof: First, from Lemma 4.1 (a) and (b), for each $x = 0,1,\ldots,N_i-1$,

$$W_i(N_i-x) = f_i(N_i-x) - \frac{x+1}{N_i-x} W_i(N_i-x-1).$$

By taking $x = \frac{N_i}{2} - 1 + z$ and after some simple computation, we have

$$W_i\left(\frac{N_i}{2} - z\right) = \frac{N_i+2-2z}{N_i+2z} f_i\left(\frac{N_i}{2} - z + 1\right) \\ - \frac{N_i+2-2z}{N_i+2z} W_i\left(\frac{N_i}{2} - z + 1\right). \tag{4.2}$$

Then, by (4.2), Lemma 4.1(c) and induction, we conclude that for each $z = 1,2,\ldots,\frac{N_i}{2}$, $W_i\left(\frac{N_i}{2} - z\right)$ can be represented as a linear function of $f_i(y)$, $y \varepsilon \{0,1,\ldots,N_i\}$.

Finally, by Lemma 4.1 (b), we also see that for each $x = \frac{N_i}{2} + 1,\ldots,N_i$, $W_i(x)$ can be represented as a linear function of $f_i(y)$, $y \varepsilon \{0,1,\ldots,N_i\}$. Hence, the proof of this theorem is completed.

By Theorem 4.1, for each $x = 0,1,\ldots,N_i$,

$$W_i(x) = \sum_{y=0}^{N_i} \beta(N_i,x,y)\ f_i(y), \tag{4.3}$$

where the coefficients $\beta(N_i,x,y)$ depend on N_i, x and y. Also, the values of $\beta(N_i,x,y)$ can be obtained from Lemma 4.1 (c) and the iterative relation (4.2).

We then define

$$W^1_{in}(x) = \sum_{y=0}^{N_i} \beta(N_i,x,y)\ f^1_{in}(y) \tag{4.4}$$

where $f^1_{in}(y)$ have been defined in (4.1).

Now, define

$$\Delta^1_{in}(\underset{\sim}{x}) = \begin{cases} W^1_{0n}(x_0)f^1_{in}(x_i) - W^1_{in}(x_i)f^1_{0n}(x_0) & \text{if } p_0 \text{ is unknown,} \\ \\ p_0 f^1_{in}(x_i) - W^1_{in}(x_i) & \text{if } p_0 \text{ is known,} \end{cases} \tag{4.5}$$

and

$$d^1_n(\underset{\sim}{x}) = \{i|\Delta^1_{in}(\underset{\sim}{x}) \leq 0\}. \tag{4.6}$$

From (4.1), it is clear that $f^1_{in}(x) \to f_i(x)$ with probability 1 as $n \to \infty$ for each $x \varepsilon \{0,1,\ldots,N_i\}$. Therefore, from (4.3) and (4.4), $W^1_{in}(x) \to W_i(x)$ with probability 1 as $n \to \infty$ for each $x \varepsilon \{0,1,\ldots,N_i\}$. Thus we have the following theorem:

Theorem 4.2. Suppose that the prior distributions $G_i(\cdot)$ are symmetrical about $p = 1/2$ and N_i are even integers for all $i = (0),1,\ldots,k$. Then, the sequence of decision rules $\{d^1_n\}^\infty_{n=1}$ is asymptotically optimal relative to the prior distribution G.

Rate of Convergence of Empirical Bayes Rules $\{d^1_n\}$

We now consider the rate of convergence of the empirical Bayes rules $\{d^1_n\}$. Following the same discussion as given in (3.4) through (3.6), and

the fact that the estimators $\{f^1_{in}(x)\}$ defined in (4.1) share the same property as that defined in (3.1), it suffices to consider the behavior of $P\{W^1_{in}(x) - W_i(x) > \delta\}$ and $P\{W^1_{in}(x) - W_i(x) < -\delta\}$ as $n \to \infty$ for some $\delta > 0$, for each $x \in \{0,1,\ldots,N_i\}$, $i = (0), 1,\ldots,k$.

From (4.3) and (4.4), for each $x \in \{0,1,\ldots,N_i\}$,

$$P\{W^1_{in}(x) - W_i(x) > \delta\} = P\left\{\sum_{y=0}^{N_i} \beta(N_i,x,y)\ [f^1_{in}(y)-f_i(y)] > \delta\right\}$$

$$\leq \sum_{y=0}^{N_i} P\{\beta(N_i,x,y)\ [f^1_{in}(y)-f_i(y)] > \delta_1\}$$

where $\delta_1 = \frac{\delta}{N_i+1}$. If $\beta(N_i,x,y) = 0$ for some $0 \leq y \leq N_i$, then $P\{\beta(N_i,x,y)[f^1_{in}(y) - f_i(y)] > \delta_1\} = 0$. So, we assume $\beta(N_i,x,y) \neq 0$. When $\beta(N_i,x,y) > 0$, then

$$P\{\beta(N_i,x,y)[f^1_{in}(y)-f_i(y)] > \delta_1\} = P\{f^1_{in}(y) - f_i(y) > \frac{1}{\beta(N_i,x,y)}\}.$$

When $\beta(N_i,x,y) < 0$, then

$$P\{\beta(N_i,x,y)[f^1_{in}(y) - f_i(y)] > \delta_1\} = P\{f^1_{in}(y) - f_i(y) < \delta_1/\beta(N_i,x,y)\}\ .$$

In either case, the problem can be reduced to considering the convergence rate of $P\{|f^1_{in}(y) - f_i(y)| > \delta_2\}$ as $n \to \infty$ for some $\delta_2 > 0$. Similarly, for the convergence rate of $P\{W^1_{in}(x) - W_i(x) < -\delta\}$ where $x \in \{0,1,\ldots,N_i\}$ and $\delta > 0$, we also get a similar result. Therefore, by applying Bernstein's

inequality and following an argument similar to that of (3.7), we conclude the following theorem:

Theorem 4.3. Let $\{d_n^1\}_{n=1}^{\infty}$ be the sequence of decision rules defined in (4.6). Then, $\{d_n^1\}_{n=1}^{\infty}$ is asymptotically optimal at least of order $\exp\{-c_2 n\}$ relative to the prior distribution G for some $c_2 > 0$.

5. Smooth Empirical Estimation of $f_i(x)$ and $W_i(x)$

In this section, we also assume that $G_i(\cdot)$ are symmetrical about $p = 1/2$ and N_i are even integers for all $i = (0), 1, \ldots, k$. In Section 4, the marginal frequency functions $f_i(x)$, $x \varepsilon \{0, 1, \ldots, N_i\}$, $i = (0), 1, \ldots, k$, are estimated in terms of the empirical frequency functions $f_{in}^1(x)$, regardless of the properties associated with the marginal function $f_i(x)$. In this section, by considering some properties related to $f_i(x)$ and $W_i(x)$, two methods for smoothing the estimators $f_{in}^1(x)$ and $W_{in}^1(x)$ are studied.

We first need the following lemmas.

Lemma 5.1. Suppose that $G_i(\cdot)$ is symmetrical about $p = 1/2$ and N_i is an even integer. Then,

(a) $$(y+1) f_i(y+1) \leq (N_i - y) f_i(y) \text{ and } \tag{5.1}$$

(b) $$W_i(y) \leq W_i(N_i - y) \tag{5.2}$$

for all $y \varepsilon \{0, 1, \ldots, N_i/2-1\}$.

Lemma 5.1 can be verified by direct computation. We omit the proof here.

Lemma 5.2. Let $U(x)$, $h(x)$ be nonnegative functions defined on $\{0,1,\ldots,N\}$, where N is an even positive integer, and satisfy

(i) $U(x) = \frac{x+1}{N-x} U(N-x-1)$ for all $x = 0,1,\ldots,N-1$.

(ii) $U(x) + U(N-x) = h(x) = h(N-x)$ for all $x = 0,1,\ldots,N$ and

(iii) $U(x) \leq U(N-x)$ for all $x = 0,1,\ldots,N/2-1$.

Then,

(iv) $(x+1)\, h(x+1) \leq (N-x)\, h(x)$ for all $x = 0,1,\ldots,N/2-1$.

Proof: Note that from (i), $(N-x)\, U(x) = (x+1)\, U(N-x-1)$. Then, by (ii), we obtain

$$(N-x)\, [h(x) - U(N-x)] = (x+1)\, [h(x+1) - U(x+1)].$$

Hence,

$$\begin{aligned}
&(N-x)\, h(x) - (x+1)\, h(x+1) \\
&= (N-x)\, U(N-x) - (x+1)\, U(x+1) \\
&\geq (N-x)\, U(x) - (x+1)\, U(x+1) && \text{(by (iii))} \\
&\geq (N-x)\, U(x) - (x+1)\, U(N-x-1) && \text{(by (iii))} \\
&= (N-x)\, [U(x) - \tfrac{x+1}{N-x}\, U(N-x-1)] \\
&= 0 \quad \text{(by (i))}.
\end{aligned}$$

Hence, the proof of this lemma is completed.

We note that conditions (i), (ii) and (iv) of Lemma 5.2 do not imply that $U(x) \le U(N-x)$ for all $x = 0,1,\ldots,N/2-1$. The following example illustrates this fact.

Example. Take $N = 4$. Let

$$U(0) = \frac{7}{120},\ U(1) = \frac{8}{120},\ U(2) = \frac{12}{120},\ U(3) = \frac{28}{120},\ U(4) = \frac{5}{120} \text{ and}$$

$$h(0) = 0.1,\ h(1) = 0.3,\ h(2) = 0.2,\ h(3) = 0.3,\ h(4) = 0.1$$

Then, conditions (i), (ii) and (iv) are satisfied but $U(4) < U(0)$.

From Lemma 5.1, the inequalities (5.1) and (5.2) are always true for all $y = 0,1,\ldots,N_i/2-1$. However, the empirical frequency functions $f^1_{in}(x)$ and the functions $W^1_{in}(x)$ do not always satisfy the above inequalities. Hence, it is reasonable to consider some smoothing of $f^1_{in}(x)$ and $W^1_{in}(x)$, which will satisfy the above inequalities. Two smoothing methods, based on $f^1_{in}(x)$ and $W^1_{in}(x)$, respectively, are given as follows.

Method 1. Smoothing Based on $f^1_{in}(x)$

Let $\Delta_1 \ge 0$ (for $\Delta_1 > 0$, Δ_1 is chosen small). Let m_1 stand for the number of times the smoothing process is carried out. Algorithmically, first $m_1 = 0$.

Step 1. $m_1 = m_1 + 1$.

For each $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, let $\varepsilon_{in}(\Delta_1, y) = \min(\Delta_1, f^1_{in}(y) + f^1_{in}(y + 1))$.

Check whether $(N_i-y)f^1_{in}(y) - (y+1)f^1_{in}(y+1) - \varepsilon_{in}(\Delta_1,y) \ge 0$ or not. If not, for $y \le \frac{N_i}{2} - 2$, let

$$a_{in}(\Delta_1,y) = [(y+1)f^1_{in}(y+1)-(N_i-y)f^1_{in}(y)+\varepsilon_{in}(\Delta_1,y)]/(N_i+1),$$

$$f^0_{in}(y) = f^0_{in}(N_i-y) = f^1_{in}(y) + a_{in}(\Delta_1,y),$$

$$f^0_{in}(y+1) = f^0_{in}(N_i-y-1) = f^1_{in}(y) - a_{in}(\Delta_1,y) \text{ and}$$

$$f^0_{in}(x) = f^1_{in}(x) \text{ for all } x \neq y,\ y+1,\ N_i-y-1,\ N_i-y.$$

For $y = \frac{N_i}{2} - 1$, let

$$a_{in}(\Delta_1,y) = [(y+1)f^1_{in}(y+1) - (N_i-y)f^1_{in}(y) + \epsilon_{in}(\Delta_1,y)] \times \frac{4}{3N_i+2},$$

$$f^0_{in}(y) = f^0_{in}(N_i-y) = f^1_{in}(y) + \frac{1}{2}a_{in}(\Delta_1,y),$$

$$f^0_{in}(y+1) = f^1_{in}(y+1) - a_{in}(\Delta_1,y) \text{ and}$$

$$f^0_{in}(x) = f^1_{in}(x) \quad \text{for all } x \neq \frac{N_i}{2} - 1,\ \frac{N_i}{2} \text{ and } \frac{N_i}{2} + 1.$$

Step 2. Let $\epsilon^0_{in}(\Delta_1,y) = \min(\Delta_1, f^0_{in}(y) + f^0_{in}(y+1))$. Check whether $(N_i-y)f^0_{in}(y) - (y+1)f^0_{in}(y+1) - \epsilon^0_{in}(\Delta_1,y) \geq 0$ for all $y = 0,1,\ldots,\frac{N_i}{2} - 1$ or not.
If yes, go to step 3.
If no, let $f^1_{in}(x) = f^0_{in}(x)$ for all $x = 0,1,\ldots,N_i$, and go to step 1.

Step 3. Define $W^0_{in}(x) = \sum_{y=0}^{N_i} \beta(N_i,x,y)f^0_{in}(y)$, $x = 0,1,\ldots,N_i$.

Remark 5.1. (1). We note that when the above smoothing procedure stops, then the smooth estimators $f^0_{in}(y)$ have the property that $(y+1)f^0_{in}(y+1) \leq (N_i-y)f^0_{in}(y)$ for all $y = 0,1,\ldots,\frac{N_i}{2} - 1$.

(2). However, it is possible that the above smoothing procedure never stops. In this situation, we can set up a maximal smoothing time to stop this procedure. When this happens, the inequality that $(y+1)f^0_{in}(y+1) \leq (N_i-y)f^0_{in}(y)$ for all $y = 0,1,\ldots,\frac{N_i}{2} - 1$, is not guaranteed.

Based on the smooth estimators $f^0_{in}(x)$ and $W^0_{in}(x)$, we define decision rules $d^0_n(\cdot)$, $n = 1,2,\ldots,$ as follows:

$$d^0_n(\underset{\sim}{x}) = \begin{cases} \{i \mid W^0_{0n}(x_0)f^0_{in}(x_i) - W^0_{in}(x_i)f^0_{0n}(x_0) \leq 0\} & \text{if } P_0 \text{ is unknown;} \\ \{i \mid P_0 f^0_{in}(x_i) - W^0_{in}(x_i) \leq 0\} & \text{if } P_0 \text{ is known.} \end{cases}$$

Method 2. Smoothing Based on $W^1_{in}(x)$

Let $\Delta_2 \geq 0$ (for $\Delta_2 > 0$, Δ_2 is chosen small). We start with a variable m_2 which stands for the number of times the smoothing carried. At first $m_2 = 0$.

Step 1. $m_2 = m_2 + 1$.

For each $y = 0, 1, \ldots, \frac{N_i}{2} - 1$, let $\delta_{in}(\Delta_2, y) = \min(\Delta_2, [W^1_{in}(y)+W^1_{in}(N_i-y)]/2)$ and $b_{in}(\Delta_2, y) = [W^1_{in}(y) - W^1_{in}(N_i - y)]/2 + \delta_{in}(\Delta_2, y)$.

Check whether $W^1_{in}(N_i-y) \geq W^1_{in}(y) + \delta_{in}(\Delta_2,y)$ or not. If not, for $y = 0$, let

$$W^*_{in}(N_i) = W^1_{in}(N_i) + c(0)b_{in}(\Delta_2,0),$$

$$W^*_{in}(0) = W^1_{in}(0) - d(0)b_{in}(\Delta_2,0),$$

$$W^*_{in}(N_i-1) = N_i W^*_{in}(0) \text{ and}$$

$$W^*_{in}(x) = W^1_{in}(x) \quad \text{for all } x \neq 0, N_i-1, N_i.$$

For $1 \le y \le \frac{N_i}{2} - 1$, let

$$W^*_{in}(N_i-y) = W^1_{in}(N_i-y) + c(y)b_{in}(\Delta_2,y),$$

$$W^*_{in}(y) = W^1_{in}(y) - d(y)b_{in}(\Delta_2,y),$$

$$W^*_{in}(N_i-y-1) = \frac{N_i-y}{y+1} W^*_{in}(y)$$

$$W^*_{in}(y-1) = \frac{y}{N_i-y+1} W^*_{in}(N_i-y) \text{ and}$$

$$W^*_{in}(x) = W^1_{in}(x) \text{ for all } x \ne y-1,\ y,\ N_i-y-1,\ N_i-y.$$

Here, $c(y) = 2(N_i-y+1)/(N_i+2)$, $d(y) = 2(y+1)/(N_i+2)$ for $y = 0,1,\ldots,\frac{N_i}{2}-1$.

Step 2. Let $\delta^*_{in}(\Delta_2,y) = \min(\Delta_2,[W^*_{in}(y) + W^*_{in}(N_i-y)]/2)$. Check whether $W^*_{in}(y) + \delta^*_{in}(\Delta_2,y) \le W^*_{in}(N_i-y)$ for all $y = 0,1,\ldots, \frac{N_i}{2} - 1$ or not.

If yes, go to step 3.

If no, let $W^1_{in}(x) = W^*_{in}(x)$ for all $x = 0,1,\ldots,N_i$ and go to step 1.

Step 3. Let $f^*_{in}(y) = W^*_{in}(y) + W^*_{in}(N_i-y)$ for all $y = 0,1,\ldots,N_i$.

Remark 5.2. (1) We note that when the above smoothing procedure stops, then the smooth estimators $W^*_{in}(y)$ satisfy that $W^*_{in}(y) \le W^*_{in}(N_i-y)$ for all $y = 0,1,\ldots,\frac{N_i}{2} - 1$, $W^*_{in}(y) = \frac{y+1}{N_i-y} W^*_{in}(N_i-y-1)$ for all $y = 0,1,\ldots,N_i-1$ and $W^*_{in}(y) + W^*_{in}(N_i-y) = f^*_{in}(y) = f^*_{in}(N_i-y)$ for all $y = 0,1,\ldots,N_i$. Then by Lemma 5.2, $(y+1)f^*_{in}(y+1) \le (N_i-y)f^*_{in}(y)$ for all $y = 0,1,\ldots,\frac{N_i}{2} - 1$. Therefore, method 2 is better than method 1 in this sense.

(2) It is also possible that the above smoothing procedure never stops. Hence, we can set up a maximal smoothing time to stop this procedure. When this happens, for the smooth estimators, the inequality properties of (5.1) and (5.2) are not guaranteed.

Based on the smooth estimators $f^*_{in}(x)$ and $W^*_{in}(x)$, we define decision rules $d^*_n(.)$, $n = 1,2,\ldots$, as follows:

$$d^*_n(\underset{\sim}{x}) = \begin{cases} \{i \mid W^*_{0n}(x_0) f^*_{in}(x_i) - W^*_{in}(x_i) f^*_{0n}(x_0) \leq 0\} & \text{if } p_0 \text{ is unknown,} \\ \{i \mid p_0 f^*_{in}(x_i) - W^*_{in}(x_i) \leq 0\} & \text{if } p_0 \text{ is known.} \end{cases}$$

6. Monte Carlo Studies

For the sequence of decision rules $\{\delta_n(\underset{\sim}{x})\}_{n=1}^{\infty}$, the conditional Bayes risk at stage n + 1 given $(\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n)$ is

$$R(G,\delta_n) = \int_\Omega \sum_{\underset{\sim}{x}\in\underset{\sim}{X}} L(\underset{\sim}{p},\delta_n(\underset{\sim}{x}))\, f(\underset{\sim}{x}|\underset{\sim}{p})\, dG(\underset{\sim}{p}).$$

To measure the performance of the sequence of decision rules $\{\delta_n(\underset{\sim}{x})\}_{n=1}^{\infty}$, computing the overall risk $r(G,\delta_n) = ER(G,\delta_n)$ is needed, where the expectation E is taken with respect to $(\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n)$. For the small sample situation, it is impossible to analytically determine such values. Therefore, Monte Carlo simulation is employed.

In this section, we have carried out some Monte Carlo studies to see the performance of the sequences of decision rules $\{d^1_n\}$, $\{d^0_n\}$ and $\{d^*_n\}$. We let, conditional on p_i, $X_i \sim B(N_i,p_i)$ where N_i are even integers for i = 0, 1 and p_0 is treated as unknown. We also assume that

$$G_i(p) = \int_0^p \frac{\Gamma(2\alpha_i+2)}{[\Gamma(\alpha_i+1)]^2} y^{\alpha_i}(1-y)^{\alpha_i}\, dy,\ i = 0,1.$$

Hence, the Bayes rule d_G is:

$$\text{Select } \pi_1 \text{ as good if } \frac{X_0+\alpha_0+1}{N_0+2\alpha_0+2} \le \frac{X_1+\alpha_1+1}{N_1+2\alpha_1+2}\ .$$

A random sample of size 50 was generated by computing from a population having $f_i(x)$ $(i=0,1)$ as probability function. For each $n = 1,2,\ldots,50$, the conditional Bayes risks $R(G,d_n^1)$, $R(G,d_n^0)$ and $R(G,d_n^*)$ were calculated. One hundred repetitions were performed. Estimates of the overall risks $r(G,d_n^1)$, $r(G,d_n^0)$ and $r(G,d_n^*)$ were obtained by averaging the associated conditional Bayes risks and the standard deviations of the estimated overall risks were also obtained based on these repeated samples.

In Tables 1-4, we consider the combinations of different N_i's and α_i's values for our decision problem. We let $\hat{r}(G,\delta_n)$ denote the average of 100 $R(G,\delta_n)$ values obtained from simulation. The standard deviation associated with $\hat{r}(G,\delta_n)$ is given in the corresponding parentheses. It is easy to see that the performances of the sequences of decision rules $\{d_n^0\}$ and $\{d_n^*\}$ are always better than that of $\{d_n^1\}$, for the cases that $(N_0, N_1, \alpha_0, \alpha_1) = (2,2,4,4)$ and $(N_0,N_1,\alpha_0,\alpha_1) = (2,2,6,6)$, both of them have the same performance. For the other $(N_0,N_1,\alpha_0,\alpha_1)$'s, the performance of $\{d_n^*\}$ is always better than that of $\{d_n^0\}$.

It is also interesting to note that in most cases, $\hat{r}(G,d_n^*)$ has the smallest standard deviation while $\hat{r}(G,d_n^1)$ has the largest standard deviation. This fact indicates that the behavior of the sequence of decision rules $\{d_n^*\}$ is more stable than the others.

TABLE 1

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 4, 4)$, $r(G) = 0.05287$

n	$\hat{r}(G, d_n^1)$	$\hat{r}(G, d_n^0)$	$\hat{r}(G, d_n^*)$
1	0.08936	0.06299	0.06299
	(0.00118)	(0.00073)	(0.00073)
2	0.08426	0.05907	0.05907
	(0.00171)	(0.00062)	(0.00062)
3	0.09188	0.05638	0.05638
	(0.00213)	(0.00053)	(0.00053)
5	0.08745	0.05452	0.05452
	(0.00227)	(0.00038)	(0.00038)
10	0.08299	0.05298	0.05298
	(0.00222)	(0.00010)	(0.00010)
15	0.07899	0.05287	0.05287
	(0.00251)	(0.00000)	(0.00000)
20	0.07767	0.05287	0.05287
	(0.00243)	(0.00000)	(0.00000)
25	0.07849	0.05287	0.05287
	(0.00234)	(0.00000)	(0.00000)
30	0.07481	0.05287	0.05287
	(0.00212)	(0.00000)	(0.00000)
35	0.07328	0.05287	0.05287
	(0.00215)	(0.00000)	(0.00000)
40	0.07215	0.05287	0.05287
	(0.00203)	(0.00000)	(0.00000)
45	0.07200	0.05287	0.05287
	(0.00213)	(0.00000)	(0.00000)
50	0.07157	0.05287	0.05287
	(0.00212)	(0.00000)	(0.00000)

TABLE 2

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (2, 2, 6, 6)$, $r(G) = 0.04896$

n	$\hat{r}(G, d_n^1)$	$\hat{r}(G, d_n^0)$	$\hat{r}(G, d_n^*)$
1	0.07640	0.05626	0.05626
	(0.00089)	(0.00056)	(0.00056)
2	0.07188	0.05284	0.05284
	(0.00126)	(0.00043)	(0.00043)
3	0.07577	0.05121	0.05121
	(0.00160)	(0.00038)	(0.00038)
5	0.07272	0.04965	0.04965
	(0.00175)	(0.00022)	(0.00022)
10	0.07098	0.04896	0.04896
	(0.00164)	(0.00000)	(0.00000)
15	0.07111	0.04896	0.04896
	(0.00189)	(0.00000)	(0.00000)
20	0.07090	0.04896	(0.04896)
	(0.00173)	(0.00000)	(0.00000)
25	0.06931	0.04896	(0.04896)
	(0.00185)	(0.00000)	(0.00000)
30	0.06904	0.04896	0.04896
	(0.00176)	(0.00000)	(0.00000)
35	0.06938	0.04896	0.04896
	(0.00181)	(0.00000)	(0.00000)
40	0.06855	0.04896	0.04896
	(0.00171)	(0.00000)	(0.00000)
45	0.06860	0.04896	0.04896
	(0.00169)	(0.00000)	(0.00000)
50	0.06681	0.04896	0.04896
	(0.00169)	(0.00000)	(0.00000)

TABLE 3

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (4, 4, 4, 4)$, $r(G) = 0.04114$

n	$\hat{r}(G, d_n^1)$	$\hat{r}(G, d_n^0)$	$\hat{r}(G, d_n^*)$
1	0.08641 (0.00030)	0.07389 (0.00123)	0.06350 (0.00123)
2	0.09043 (0.00168)	0.06071 (0.00164)	0.05480 (0.00116)
3	0.08626 (0.00069)	0.06396 (0.00131)	0.05026 (0.00092)
5	0.08595 (0.00174)	0.05851 (0.00156)	0.04613 (0.00069)
10	0.08479 (0.00145)	0.05546 (0.00147)	0.04274 (0.00035)
15	0.07992 (0.00161)	0.05596 (0.00147)	0.04189 (0.00016)
20	0.07994 (0.00156)	0.05536 (0.00145)	0.04148 (0.00010)
25	0.07514 (0.00185)	0.05358 (0.00133)	0.04155 (0.00011)
30	0.07674 (0.00157)	0.05446 (0.00141)	0.04144 (0.00011)
35	0.07458 (0.00156)	0.05449 (0.00139)	0.04144 (0.00010)
40	0.07024 (0.00158)	0.05312 (0.00130)	0.04168 (0.00013)
45	0.06855 (0.00145)	0.05267 (0.00126)	0.04158 (0.00013)
50	0.06749 (0.00152)	0.05187 (0.00119)	0.04158 (0.00012)

TABLE 4

Simulation results for the comparative performance of sequences of empirical Bayes rules $\{d_n^1\}$, $\{d_n^0\}$ and $\{d_n^*\}$.

$(N_0, N_1, \alpha_0, \alpha_1) = (4, 4, 6, 6)$, $r(G) = 0.03970$

n	$\hat{r}(G, d_n^1)$	$\hat{r}(G, d_n^0)$	$\hat{r}(G, d_n^*)$
1	0.07415 (0.00020)	0.06354 (0.00088)	0.05607 (0.00087)
2	0.07609 (0.00117)	0.05526 (0.00119)	0.04950 (0.00073)
3	0.07333 (0.00057)	0.05737 (0.00104)	0.04669 (0.00066)
5	0.07158 (0.00110)	0.05513 (0.00108)	0.04275 (0.00051)
10	0.07102 (0.00132)	0.05097 (0.00103)	0.04079 (0.00024)
15	0.06987 (0.00102)	0.05202 (0.00106)	0.04006 (0.00009)
20	0.06907 (0.00122)	0.05035 (0.00105)	0.04001 (0.00009)
25	0.06632 (0.00121)	0.04973 (0.00100)	0.04004 (0.00010)
30	0.06748 (0.00118)	0.05067 (0.00099)	0.04004 (0.00010)
35	0.06669 (0.00136)	0.05024 (0.00101)	0.03997 (0.00009)
40	0.06712 (0.00118)	0.05034 (0.00098)	0.04004 (0.00009)
45	0.06626 (0.00129)	0.05110 (0.00101)	0.03987 (0.00006)
50	0.06373 (0.00126)	0.05042 (0.00102)	0.03992 (0.00008)

References

Deely, J. J. (1965). Multiple decision procedures from an empirical Bayes approach. Ph.D. Thesis (Mimeo, Ser. No. 45), Dept. of Statist., Purdue Univ., West Lafayette, IN.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. J. Statist. Plan. Infer., 8, 87-101.

Huang, W. T. (1975). Bayes approach to a problem of partitioning k normal populations. Bull. Inst. Math. Acad. Sinica, 3, 87-97.

Ibragimov, I. A. and Linnik, Yu. V. (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff Publishing Groningen.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Statist. Prob., 155-163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 35, 1-20.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypothesis. Ann. Math. Statist., 34, 1370-1385.

Singh, A. K. (1977). On slippage tests and multiple decision (selection and ranking) procedures. Ph.D. Thesis. (Mimeo. Ser. No. 494), Dept. of Statist., Purdue University, West Lafayette, IN.

Van Ryzin, J. (1970). On some nonparametric empirical Bayes multiple decision problems. Proc. First Internat. Symp. Nonparametric Techniques in Statistic Inference, (ed. M. L. Puri), 585-603.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist., 5, 172-181.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #84-37	2. GOVT ACCESSION NO.: AD-A147614
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Empirical Bayes Rules for Selecting Good Binomial Populations	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: Technical Report #84-37
7. AUTHOR(s): Shanti S. Gupta and Ta Chen Liang	8. CONTRACT OR GRANT NUMBER(s): N00014-84-C-0167
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, D.C.	12. REPORT DATE: October 1984
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayes Rules; Empirical Bayes Rules; Prior Distribution; Asymptotically Optimal; Rate of Convergence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper deals with the problem of selecting good binomial populations compared with a standard or a control through the empirical Bayes approach. Two cases have been studied: one is that the prior distribution is completely unknown and the other is that the prior distribution is symmetrical about $p = 1/2$, but its form is still unknown. In each case, empirical Bayes rules are derived and the rate of convergence of corresponding empirical Bayes rules is also studied. In each case, it is found that the order of the rate of convergence is $O(\exp(-c_i n))$ for some $c_i > 0$, $i = 1,2$.

END

FILMED

12-84

DTIC